Powered by Futurense

Foundations of Statistics & Probability

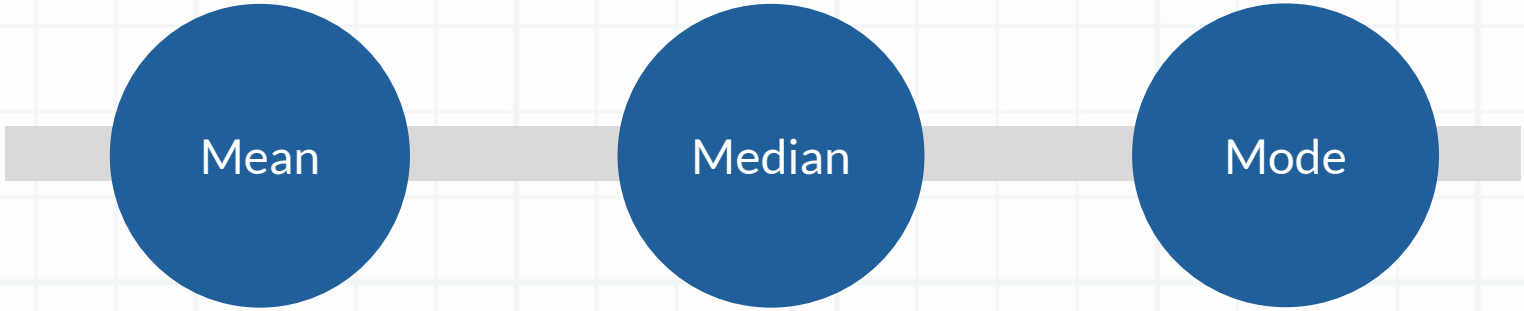
Module 01 Introduction to Statistics

Descriptive Statistics Contd.
Measures of Central Tendency

Dr. Rajlaxmi Chouhan
Associate Professor, Department of Electrical Engineering
Head, Center for Education & Technology for Education
IIT Jodhpur

Measures of Central Tendency

Measures of central tendency are statistical tools used to summarize a dataset by identifying the central or typical value within it.



Mean

Mean represents the **average** value of all the items, i.e. sum of all individual items / total number of items

Example

A group of 20 college students in India is surveyed about their monthly mobile data usage.

The data (in GB) is as follows:

Dataset
5, 8, 10, 10, 12, 12, 12, 15, 15, 15,
18, 20, 20, 25, 30, 35, 40, 50, 60, 100

Mean

***N* = 20**

Dataset
5, 8, 10, 10, 12, 12, 12, 15, 15, 15, 18, 20, 20, 25, 30, 35, 40, 50, 60, 100

Mean = Sum / N

Sum

= 5 + 8 + 10 + 10 + 12 + 12 + 12 + 15 + 15 + 15 + 18 + 20
+ 20 + 25 + 30 + 35 + 40 + 50 + 60 + 100

= 512

Mean = 512 / 20 = 25.6 GB

Average consumption of mobile data

Median

1. ***Arrange the data in ascending order.***

   5, 8, 10, 10, 12, 12, 12, 15, 15, 15, 18, 20, 20, 25, 30, 35, 40, 50, 60, 100

2. *Find the **middle values** (**N/2**th and **N/2 + 1**st for an even dataset)*

   5, 8, 10, 10, 12, 12, 12, 15, 15, **15, 18**, 20, 20, 25, 30, 35, 40, 50, 60, 100

3. **Median = (15 + 18) / 2 = 16.5 GB**

   *The median reflects the **middle range of data usage (representative value).***

Mode

Most frequent value

5, 8, 10, 10, 12, 12, 12, 15, 15, 15, 18, 20, 20, 25, 30, 35, 40, 50, 60, 100

Mode = 12 and 15 (each occurs 3 times)

Inferences

5, 8, 10, 10, 12, 12, 12, 15, 15, 15, 18, 20, 20, 25, 30, 35, 40, 50, 60, 100

Mean (25.6 GB)

Provides the overall average, **incorporating all values in the dataset**.

Prone to outliers

Median (16.5 GB)

Reflects the middle range of data usage, unaffected by high outliers (like 100 GB) and **better represents the majority of students**

Not prone to outliers

Mode (12 GB, 15 GB)

Shows the most common data consumption values. **'most popular choices'**

Inferences

5, 8, 10, 10, 12, 12, 12, 15, 15, 15, 18, 20, 20, 25, 30, 35, 40, 50, 60, 100

Mean (25.6 GB)

Provides the overall average, **incorporating all values in the dataset**.

Prone to outliers

Median (16.5 GB)

Reflects the middle range of data usage, unaffected by high outliers (like 100 GB100GB) and **better represents the majority of students**

Not prone to outliers

Mode (12 GB, 15 GB)

Shows the most common data consumption values. **'most popular choices'**

Which of these measures appear to be the better central measure for this particular dataset?

Significance & Uses

Mean (25.6 GB)

Provides the overall average, **incorporating all values in the dataset**.

Prone to outliers

Where is the *mean* useful?

Scenario 1: Evaluating Overall Usage

Telecom companies might use the **mean** to determine the average data demand across their customer base to plan infrastructure upgrades or pricing strategies.

Scenario 2: Comparing Groups

If comparing data usage between urban and rural students, the mean gives a quick overview of which group uses more data on average.

Scenario 3: Budgeting Resources

A university planning to provide mobile data subsidies could use the mean to estimate the total amount of data required for all students.

Significance & Uses

Mean (25.6 GB)

Provides the overall average, **incorporating all values in the dataset**.

Prone to outliers

Where can the *mean* be misleading?

Individual-level insights

The mean (25.6 GB) suggests that most students use around 26 GB of data, but the median (16.5 GB) shows typical usage.

Significance & Uses

Mean (25.6 GB)

Provides the overall average, **incorporating all values in the dataset**.

Prone to outliers

Where can the *mean* be misleading?

Individual-level insights

The mean (25.6 GB) suggests that most students use around 26 GB of data. Use the **median (16.5 GB)** shows typical usage.

Decision-Making for Popular Plans

If a telecom company relies on the mean to design data plans, they might set a high base plan (e.g., 30 GB) that does not match most students' needs.

Use the **mode (12 GB, 15 GB)** to identify the most popular plans instead.

Recap

Significance of the Mean, Median and Mode

Choosing the right measure of central tendency

Use mean when overall averages are important.

Use median when typical values matter in skewed datasets.

Use mode for identifying common trends or popular choices.

Coming up next...

Descriptive Statistics Contd.

Sample mean and variance calculations